A Flexible Multiple Description Coding Framework for Adaptive Peer-to-Peer Video Streaming

Emrah Akyol, A. Murat Tekalp, *Fellow, IEEE*, and M. Reha Civanlar, *Fellow, IEEE*

***Abstract*—Efficient peer-to-peer (P2P) video streaming is a challenging task due to time-varying nature of both the number of available peers and network/channel conditions. This paper proposes a novel adaptive P2P video streaming system, which features: i) a new flexible multiple-description coding (F-MDC) framework, such that the number of base and enhancement descriptions, and the rate and redundancy level of each description can be adapted on the fly (by only post-processing of the encoded bitstream), and ii) a new adaptive TCP-friendly rate-controlled (TFRC), on-demand, many-to-one P2P video streaming solution based on the proposed F-MDC framework. We extend the highly scalable video coder [17], [18] to MDC within the proposed F-MDC framework. Optimization of the design parameters of the proposed F-MDC method is discussed within the context of the proposed adaptive P2P streaming solution, where the number and quality of available streaming peers/paths are *a priori* unknown and vary in time. Experimental results, by means of NS-2 network simulation of a P2P video streaming system, show that adaptation of the number, type of descriptions and the rate and redundancy level of each description according to network conditions yields significantly superior performance when compared to other scalable MDC schemes using a fixed number of descriptions/layers with fixed rate and redundancy level.**

***Index Terms*—Adaptive P2P streaming, multiple description coding, rate-distortion optimization, scalable video coding, TCP-friendly rate-controlled (TFRC).**

I. Introduction

MULTIPLE description coding (MDC) addresses the problem of encoding source information using more than one independently decodable and complementary bitstreams, which, when combined, can provide the highest level of quality and when used independently, can still provide an acceptable level of quality. This is made possible by introducing some redundancy in each description, which will be discarded if all streams are received [1]. It is well known that MDC can provide robust video communication over unreliable networks, such as the Internet, when combined with path/server diversity at the cost of reduced compression efficiency [2]. There has been a significant amount of work on multiple description video coding [2]–[8]. Notably, in [3] motion estimation across descriptions, called motion compensated multiple description coding (MC-MDC) is utilized. Unbalanced multiple description video coding is studied in [4], where descriptions do not have identical rates, i.e., one or more descriptions are coded at a lower bitrate than others. It is shown that unbalanced MD is especially useful for Internet streaming where paths with different bandwidths are common. An MDC framework based on Discrete Wavelet Transform (DWT) is proposed in [5] allowing redundancy adaptation to varying wireless channel conditions. However, that approach does not allow post-encoding adaptation, which is paramount for content distribution since peers with content have limited storage capacity. In order to make post-encoding adaptation possible, MDC should be employed with a scalable video coding scheme.

One example of scalable MDC is based on motion compensated temporal filtering, where high frequency frames are grouped into two descriptions and missing frames are estimated using motion vectors in the two descriptions [7]. It is reported to outperform existing nonscalable MD video coders in compression efficiency, while providing flexible rate allocation and redundancy control, although its performance is degraded under high motion since estimating missing frames then becomes difficult. In [8], forward error correction (FEC) is used with MDC. FEC-MDC unequally protects a progressive bitstream with erasure channel codes according to the importance of bitstream segments. Every description includes a protected version of the most important layer, then half of the next important layer, and so on. However, this requires a significantly high number of descriptions (such as $N = 16$ or $N = 32$) to allocate redundancy effectively; such a high number of descriptions will deteriorate the compression efficiency especially at low packet loss rates. Also, determining the optimal FEC allocation according to the varying channel conditions on the fly is difficult. In [9], a network-adaptive, unbalanced MDC method is proposed by grouping 3-D SPIHT coefficients into two unequal groups, and applying unequal error protection to bit planes. The amount of FEC is adapted to packet loss rates, and the number of wavelet coefficients to code for each sender is allocated according to the estimated TFRC rate. We note that, none of the available MDC methods addresses post encoding adaptation of the number of descriptions/layers and the amount of redundancy in descriptions/layers according to the network conditions.

Manuscript received November 2, 2006; revised May 4, 2007. This work was done when all authors were at Koç University, Istanbul, Turkey. A preliminary version of this work was presented in part at the IEEE International Conference on Image Processingin 2006 and the European Signal Processing Conference in 2005. This work was supported by the European Commission within FP6, under the Network of Excellence Grant 511568 with acronym 3DTV. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Adriana Dumitras.

E. Akyol is with the Department of Electrical Engineering, Henry Samuel School of Engineering and Applied Science, University of California, Los Angeles, CA 90095-1594, USA (e-mail: eakyol@ee.ucla.edu).

A. M. Tekalp is with the Department of Electrical and Computer Engineering, College of Engineering, Koç University, 34450 Sariyer, Istanbul, Turkey (e-mail: mtekalp@ku.edu.tr).

M. R. Civanlar is with DoCoMo USA Labs, Palo Alto, CA 94304-1201 USA (e-mail: rcivanlar@docomolabs-usa.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2007.901527

On-demand P2P video streaming has recently gained interest [14]–[16]. In on-demand P2P streaming, there is no need for dedicated edge servers to store and distribute video, instead peers who store the requested coded video starts streaming to the requesting peer when the request occurs. The requesting peer can coordinate the peer selection and streaming from multiple peers to avoid a central coordinator or specific tree structures unlike multicasting. However, P2P streaming has some challenging issues that need to be addressed.

i) *Peer Query and Selection*: Selecting the optimal peer(s) for streaming is a difficult task because of the heterogeneity of sending peer conditions. Round Trip Times (RTT) and upstream transmission rates may vary from peer to peer. Also, some of the peers may actually share a link in their path to the receiver: a situation not easily detectable. Using disjoint paths is important since any loss on the shared link effects both of the streams transmitted over two paths, rendering multipath streaming ineffective. Therefore, to utilize path diversity fully, the receiver may need to monitor the correlation of packet losses between all path pairs.

ii) *Packet Losses*: During video transmission, any sender peer may turn off, a link may be broken or packets may be lost or delayed due to competing TCP traffic. Because of the stringent delay constraints coupled with possibly high RTT values [19] lost packets may not be retransmitted in time. The proposed F-MDC can be a remedy to this problem, since the decoder can generate video with graceful degradation from the received packets under packet loss.

iii) *Low and Heterogeneous Upstream Rates*: Usually the upstream rate of an individual sending peer is much lower than the downstream capacity of the receiving peer, necessitating some kind of distributed P2P streaming. Also, peers may be connected to the Internet via different speed connections; hence, the receiving peer should manage the heterogeneous upstream capacities of selected sending peers by a rate allocation algorithm to minimize the overall distortion. To this effect, a flexible unbalanced MDC, such as the proposed F-MDC, that allows any rate partition post encoding, is needed. Rate allocation to sending peers should be performed at the receiver, since only the receiver knows the statistics of each path, and sending peers may not be willing to waste computational resources for this purpose. Moreover, there should be a reliable mechanism to send rate and packet partitioning information (control packets) from receiver to senders.

iv) *Time-Varying Network Conditions*: In P2P systems, it is common that packet loss rates and upstream capacities vary due to external traffic and/or peers may tune out unpredictably. Analysis of P2P systems shows that around 60% of peers remain active less than 10 minutes, each time they join the system [19]. Hence, the MDC algorithm should allow post-encoding adaptation according network conditions; especially the number of descriptions should be flexible.

v) *Competing TCP Traffic*: It is highly likely that there is competing TCP traffic between the sender and receiver, or on subsections of the path between them created by others. Clearly, streaming traffic should not suppress the competing TCP traffic while allocating the necessary bandwidth for streaming. To this effect, TCP Friendly Rate Control (TFRC) should be used to calculate sender bit-rates to achieve a fair distribution of the bandwidth between TCP data and video.

In [20], a general framework for receiver-driven, simultaneous distributed streaming is proposed, where the receiver coordinates the packet transmissions from each sender. A rate allocation algorithm is executed at the receiver for fair distribution of the total receiver bandwidth among heterogeneous senders with different rate and channel characteristics. Also, a packet partition algorithm running at the senders is proposed to ensure that each packet is sent only once. In [13], an adaptive layered streaming framework is proposed for P2P on-demand video streaming. A receiver-driven coordination framework is proposed with congestion control of layered streaming from multiple sender peers. In [14], P2P streaming using FEC-MDC is proposed, where each peer stores only certain descriptions coded at some fixed rate. If one serving peer fails, system searches for another peer containing the same descriptions. Using layered coding in place of multiple descriptions is also analyzed and concluded that layered coding outperforms MDC when replacement time of down peers with new ones is relatively small [15].

The main contribution of this paper is two-fold: First, we propose a novel flexible MDC (F-MDC) framework, within which we extend the highly scalable video encoder of [17], [18] to MDC in order to provide efficient adaptation of P2P streaming to the available number of peers and time-varying network conditions on-the-fly, post encoding. Second, a P2P video streaming system is designed using the proposed F-MDC method, varying: i) the number of base and enhancement descriptions, ii) the redundancy of each individual description, iii) the rate of each description/layer, iv) rate allocation among descriptions, i.e., balanced/unbalanced MDC or base/enhancement descriptions. Section II introduces the proposed F-MDC framework and its application to scalable video coder[17], [18] in detail. Section III presents the adaptive on-demand P2P streaming system. Comparative results are provided in Section IV to demonstrate the performance of the proposed F-MDC and P2P streaming system. Finally, Section V presents discussions and conclusions.

II. Flexible Multiple Description Coding Framework

We start by introducing a new flexible multiple description coding (F-MDC) framework in Section II-A. In Section II-B, we review the basics of a JPEG-2000-based "T + 2D" scalable video coding scheme, which is then extended to F-MDC in Section II-C. Determination of the design parameters for the proposed F-MDC scheme is discussed in Section II-D. The summary of the variables used in the paper is listed in Table I.

A. Proposed F-MDC Framework

The proposed F-MDC framework requires a scalable coded video bitstream, composed of C independent substreams. An example of such a codec is the JPEG-2000 based "T + 2D"

TABLE I
SUMMARY OF THE VARIABLES USED

Symbol	*Description*	*Symbol*	*Description*
N_B	number of base descriptions	N_E	number of enhancement descriptions
λ_L^B	low slope for base descriptions	λ_L^E	low slope for enhancement descriptions
λ_H^B	high slope for base descriptions	λ_H^E	high slope for enhancement descriptions
C_m^B	assignment vector for base description m	C_m^E	assignment vector for enh. description m
λ_S	starting rate for enh. descriptions	$S(\lambda)$	set of CBs truncated at λ
$d_i(\lambda)$	distortion of CB i truncated at slope λ	$r_i(\lambda)$	rate of CB i truncated at slope λ
$p_i(\lambda)$	probability of constructing CB i at λ	p_{iBL}^k	loss rate of k^{th} path for CB i at λ_L^B
C	number of CBs in a GOP	L	number of layers (set at encoding)

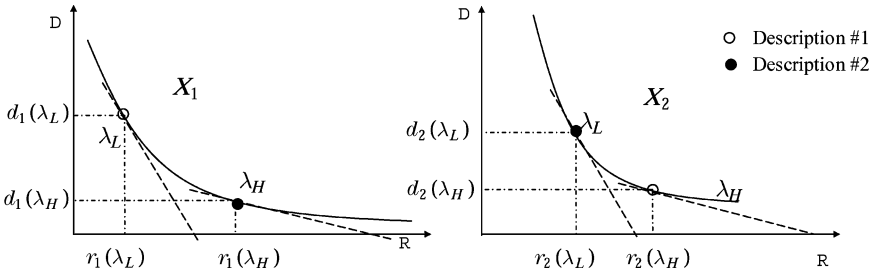


Fig. 1. Illustration of the flexible MDC for two descriptions with two independent substreams.

codec that is reviewed in Section II-B, where bitstreams for each codeblock form independent substreams. Assume that every substream can be truncated and decoded at L different rate points, where $\{r_i(\lambda_1),\ldots,r_i(\lambda_L)\}$ denote the rate points for substream i, such that L total rates $\Re = \{R_1,\ldots,R_L\}$ and rate distortion slopes $\lambda = \{\lambda_1,\ldots,\lambda_L\}$ can be obtained, where

$$R_j = \sum_{i=1}^{C} r_i(\lambda_j) \quad 1 \le j \le L. \tag{1}$$

The optimal truncation points for each substream to match a total rate can be found by the well-known Lagrange multiplier based rate-distortion optimization algorithm, given the individual rate distortion functions for each substream [24]. Then, one can generate any number of multiple descriptions by combining substreams truncated at different points, which correspond to total rates of C substreams $(R^1,\ldots,R^C)$, $1 \le i \le C$. The proposed F-MDC with optimal bit allocation can be formally defined by the following propositions.

Proposition I: If the total rate is allocated among all substreams optimally, then all substreams should have the same rate-distortion slope, provided that the total rate is high enough such that each substream can be encoded with nonzero bits, i.e., $\lambda_j^i = \lambda_j$, $1 \le i \le C$.

Proof: The proof is a direct extension of Theorem-1 of [22], assuming a convex rate-distortion function for every substream. If the total rate spent is smaller than available, optimal bit allocation can be reached by water-filling [12]. Let $S(\lambda_j)$ denote the set of substreams at a rate distortion slope λ_j. Then, an F-MDC can be realized by combining at least two different sets of substreams encoded at total rates (slopes) λ_j and λ_m, i.e., by combining $S(\lambda_j)$, and $S(\lambda_m)$.

Proposition II: Any number, $N \le C$, of multiple descriptions can be generated by combining $S(\lambda_j)$ and $S(\lambda_m)$.

Proof: A possible way of generating N multiple descriptions is by selecting one of the substreams in $S(\lambda_j)$ and remaining $N-1$ from $S(\lambda_m)$, $j \ne m$, $1 \le j,m \le L$. The simple case of two substreams and two descriptions is shown in Fig. 1. If only one of the descriptions is available one substream is decoded at slope $\lambda_j(\lambda_H = \lambda_j)$ and the other one is at $\lambda_m(\lambda_L = \lambda_m)$. If both descriptions are received, both substreams are decoded at the higher slope, i.e., λ_j.

In the following, we present the application of the proposed F-MDC framework to a highly scalable "T+2D" wavelet video codec, which uses JPEG2000 for coding motion compensated filtered frames [17], [18]. This scalable video codec fits the requirements of the proposed framework, since independently coded substreams are generated by JPEG2000 for each codeblock. Note that, independently, a similar MDC scheme competable with JPEG2000 decoder is presented in [33].

B. Basics of JPEG2000-Based "$T+2D$" Video Coding

In "T + 2D" wavelet video coding, depicted in Fig. 2, first, motion compensated temporal filtering (MCTF) is performed along the temporal direction to efficiently de-correlate frames within a GOP. Then, all filtered frames (i.e., temporal subbands) are encoded using JPEG-2000 coder. It is well known that this scheme provides state of the art compression efficiency, except possibly for very low bitrates, while providing embedded SNR, temporal and spatial scalability [17], [18]. We note that

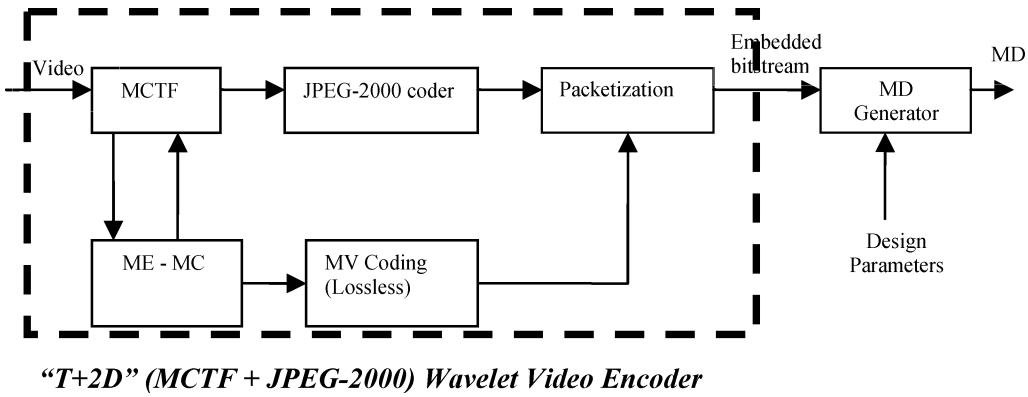


Fig. 2. General structure of the proposed MDC scheme, using the "T + 2D" wavelet video coder.

the code-block structure of JPEG-2000 with open loop motion compensation provides independently encoded substreams.

In JPEG-2000, every spatial subband is divided into small nonoverlapping block of samples called code-blocks and each code-block is encoded independent of others. Since each code-block is coded using bitplane coding, any code-block can be truncated at several points, post encoding. For every truncation point, r_i, contribution of each block to the total distortion is calculated and code-block is truncated according to its contribution to the overall distortion in rate distortion optimized manner [23]. Specifically, every truncation point correspond to some distortion-rate slope, λ. The set of optimal truncation points $\{r_i(\lambda)\}$ that minimize

$$J(\lambda) = D(\lambda) + \lambda R(\lambda) = \sum_{i=1}^{C} d_i(\lambda) + \lambda r_i(\lambda) \qquad (2)$$

for some λ is found by the post compression RD optimization algorithm explained in [21]. These points are optimal in the sense that the total distortion cannot be reduced without increasing the total rate, $R(\lambda) = \sum_{i=1}^{C} r_i(\lambda)$. Hence, there is a one-to-one mapping between the total rates and distortion-rate slopes, when the truncation points are determined in a rate-distortion optimal manner.

Ignoring the temporal drift, which is mitigated by the open loop MCTF structure, and assuming an additive distortion metric, the total distortion (mean squared error, MSE) can be written as weighted sum of code-block distortions as

$$D(\lambda) = \sum_{i=1}^{C} w_t^i w_s^i \sum_{k=1}^{K} \left(x_i(k) - Q\left(x_i(k), \lambda\right)\right)^2 \qquad (3)$$

where w_t^i and w_s^i denote temporal and spatial weights defined as the magnitude of the spatial and temporal wavelet synthesis filter coefficients [23], $x^i(k)$ and $Q(x_i(k), \lambda)$ represent the original and quantized spatio-temporal subband samples, and K is the number of coefficients in one code-block, (typically code-block size is 16×16, hence $K = 256$). Here, we neglect the nonorthogonal nature of the filters (i.e., we assume spatial and temporal filters are orthogonal such that the distortion is equivalent in the spatial and transform domain), and the correlation between the quantization errors.

The F-MDC method presented in Section II-C makes use of the code-block structure of the JPEG-2000 codec, where every code-block is independently encoded; hence, we have independently coded substreams. In JPEG-2000 syntax, the number of bits contributed from each code-block to the overall layer bitstream ($r_i(\lambda)$, for i^{th} code-block) are embedded in the packet[1] header and the corresponding distortion-rate slopes (λ) are recorded in the optional COM segment of the main header for fast post compression rate-distortion optimized truncation [23]. Hence, JPEG-2000 bitstream already contains rate-distortion information of every code-block for each layer in its packet headers.

C. Application of the Proposed F-MDC Framework to JPEG-2000-based "$T + 2D$" Video Coding

We propose to generate multiple descriptions by only post processing of a single embedded scalable video bitstream as shown in Fig. 2. Since bitplane representations of every code-block can be truncated at any point independent of other code-blocks, we first extract each codeblock at two truncation points: one corresponding to a high rate (higher slope λ_H^B) and one corresponding to a low rate (lower slope λ_L^B) to generate two different SNR quality versions of each codeblock. Then, N_B base descriptions can be generated as different combinations of these two different quality versions of code-blocks as shown in Fig. 3. Since both of these two different quality codeblocks are formed by starting from the most significant bitplanes (MSB), they can be independently decoded. Furthermore, the remaining bitplanes of each code block can also be used to generate multiple enhancement descriptions. Clearly, the enhancement descriptions require availability of the base descriptions for decodability. In the following, we describe how to generate N_B base descriptions and N_E enhancement descriptions in more detail.

1) Generation of Base Descriptions: Base descriptions are formed by various combinations of low and high rate code-blocks. For example, N_B base descriptions can be generated by including one code-block truncated at the high rate out of every N_B, and remaining $N_B - 1$ code blocks at the low rate for each description. The lowest frequency code-blocks in both temporal and spatial domains are coded at the high rate in all

[1]In JPEG-2000 terminology, a packet is a collection of coded code-blocks from the same resolution and the same layer.

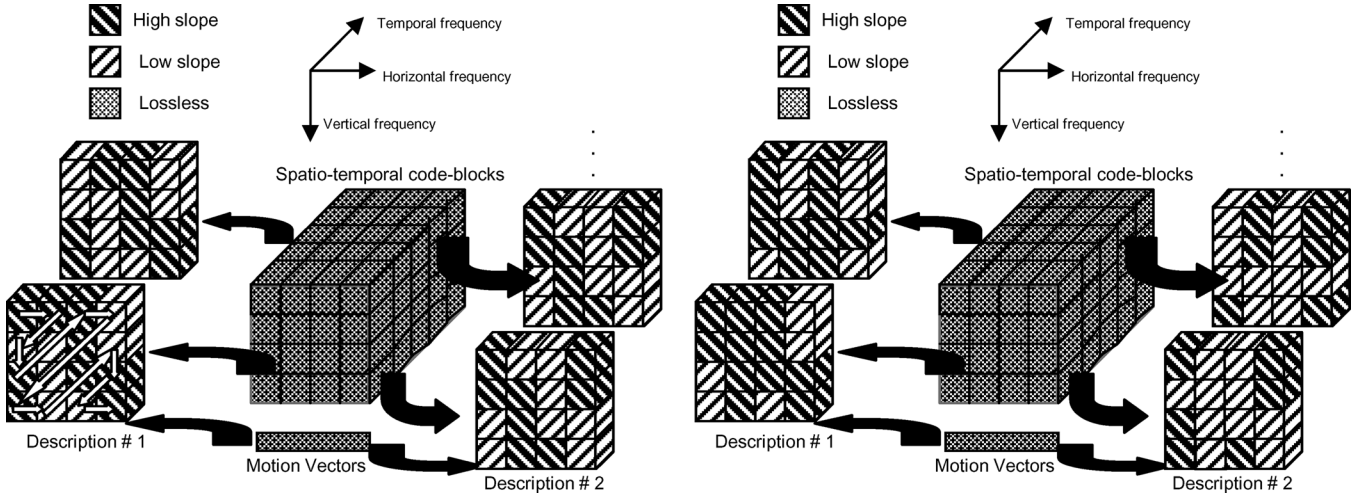


Fig. 3. Balanced and unbalanced MDC schemes generated from a single scalable bitstream. (a) balanced MDC (on the left) and (b) unbalanced MDC (on the right).

descriptions, since they affect the visual quality more than the other code-blocks. The case of $N_B = 2$ is depicted in Fig. 3, where the ordering of code-blocks follows a zig-zag scan order. We can see that description 1 has high-low-high-... rate ordered codeblocks along the zigzag scan order, whereas description 2 has low-high-low-... rate ordered codeblocks after the lowest frequency codeblock which is coded at the high rate in both descriptions. The codeblock truncation points $\{r_i(\lambda_H^B), r_i(\lambda_L^B)\}$ that correspond to λ_H^B and λ_L^B can be determined by rate-distortion optimization for each block as in EBCOT [21]. All overhead including motion vectors is coded lossless for every description. In the decoder side, if both descriptions are received, only code-blocks which are truncated at the high rate are used. On the other hand, if only one description is received, still an acceptable video quality can be achieved with some code-blocks decoded at the low rate. Since the code-blocks at different rates are extracted from a single embedded video bitstream, description rates as well as the number of descriptions generated by various combinations of them for each code-block are totally flexible, and can be varied on-the-fly post encoding. Fig. 3 illustrates a way of generating two balanced or unbalanced multiple descriptions by manipulating a scalable bitstream.

The proposed framework allows generation of both balanced and unbalanced base descriptions. Unbalanced descriptions can be generated by a combination of unequal amount of code-blocks from $S(\lambda_H^B)$ and $S(\lambda_L^B)$ in different descriptions. For example, Fig. 3(b) shows the case of $N_B = 2$, where two out of every three code-blocks is extracted at λ_H^B in description 1, and one out of three is extracted at λ_H^B in description 2.

2) Enhancement Descriptions: Multiple enhancement descriptions are generated from the remaining bitplanes (not used in the base descriptions) by specifying a starting layer, λ_S, which can be either λ_L^B or λ_H^B, and a low λ_L^E and a high layer λ_H^E using an approach similar to the one used for generating base descriptions. The decision to set $\lambda_S = \lambda_L^B$ or $\lambda_S = \lambda_H^B$ is made based on the R-D optimization. A particular example for setting λ_S in a P2P streaming application is described in Section III-C.

In summary, the process of generation of multiple base and enhancement descriptions can be completely specified in terms of the following design parameters.

1) Number of base and enhancement descriptions, N_B and N_E: Every base description has some code-blocks extracted at the high rate and others at the low rate, where the number of the code-blocks extracted at the high rate decreases as the number of descriptions increase. The number of enhancement layers can also be adjusted post encoding, similar to base descriptions.
2) High and low RD slopes to generate base descriptions (λ_H^B, λ_L^B): As λ_H^B. increases; λ_L^B should decrease in order to maintain a fixed average rate for the description.
3) Assignment of $S(\lambda_H^B)$ and $S(\lambda_L^B)$ to base descriptions, $C_m^B = [c_m^1, \ldots, c_m^C]$, $1 \leq m \leq N_B$: The vector C_m^B $(1 \times C)$ specifies the code-blocks truncated at λ_H^B and λ_L^B in description m, where C is the number of code-blocks in one GOP. The i^{th} element, $c_m^i = 1$, indicates that i^{th} code-block will be truncated at λ_H^B in description m, and $c_m^i = 0$ denotes it will be truncated at λ_L^B. Since C_m^B. needs to be sent from the receiver to all senders for every GOP, it should be expressed in the minimum form possible. To this effect, we assume that the codeblock pattern repeats periodically with a period of C' code-blocks, and a new vector $C_m^{B\prime}$ of length $C'(C' \ll C)$ is defined such that $C_m^{B\prime}(j(\text{mod } C')) = C_m^B(j)$ for $1 \leq j \leq N_B$. At the receiver side, the vector C_m^B can be easily obtained from $C_m^{B\prime}$, which is more compact to send. Also, this restriction decreases the complexity of the search for C_m^B, while sacrificing from optimality. We note that, for the base descriptions, the lowest spatio-temporal frequency is represented at the high layer without any consideration to the value of the element of the vector C_m^B for that code-block.
4) Specification of enhancement descriptions (λ_S, λ_H^E, λ_L^E, C_m^E): The starting rate (λ_S) shall be either λ_H^B or λ_L^B. Since enhancement layers are also sent as multiple descriptions, we also need to specify high/low slopes (λ_H^B, λ_L^B) for enhancement descriptions. Fig. 4 illustrates λ_S, λ_H^E, λ_L^E for

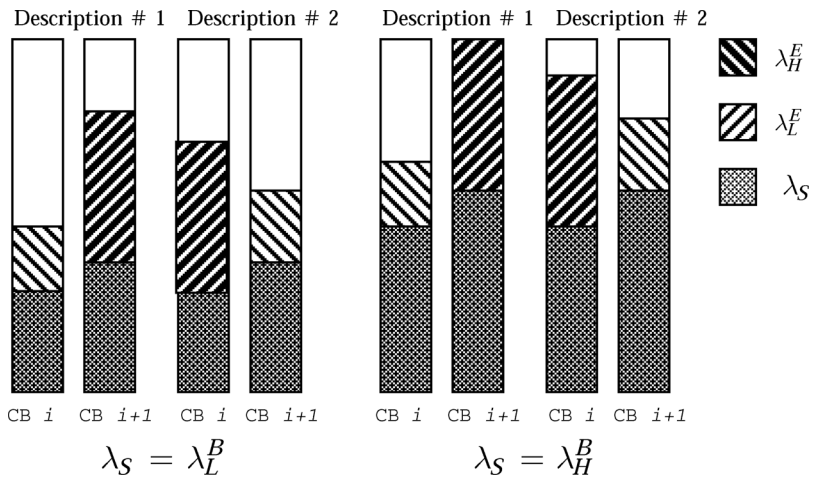


Fig. 4. Enhancement descriptions for $N_E = 2$.

the case of two code-blocks and two enhancement descriptions. We also need to send the assignment vector C_E^m, $1 \leq m \leq N_E$ for enhancement descriptions which is specified as described in step 3.

We note that the design parameters can be adjusted on the fly, post encoding. By changing the design parameters, the rate and redundancy of each description, as well as the number and the type (balanced/unbalanced or base/enhancement) of descriptions can be adapted according to transmission/network conditions on the fly, without re-encoding. Comparative results on the compression efficiency of the proposed F-MDC method, compared to other methods such as MD-MCTF [7] and MD-MDC [2], are presented in Section IV-A.

D. Determination of the Design Parameters for the Proposed F-MDC

1) Estimation of Quantization Distortion Using Hint Tracks: In JPEG-2000, the rate of each code-block for each layer is recorded in the packet header. This auxiliary information in packet headers can be packed into one group (i.e., packed packet header in JPEG-2000), encoded with tag tree coding [23] and can be sent to the receiver. The receiver can use this code-block length and layer rate-distortion slope information to deduce rate-distortion. Using packet headers in rate-distortion computation requires no extra rate since receiver already needs to know the code-block lengths for each layer to form packets contents and in decoding code-block headers. Receiver uses this pre-fetched rate-distortion information in rate-redundancy allocation before requests from senders and also in decoding the encoded code-blocks. The distortion of one code-block of up to layer k can be estimated using code-block lengths and layer rate-distortion slopes by the expression

$$d_i(\lambda_k) = d_i(\lambda_L) + \sum_{j=L}^{k} \lambda_j \left[r_i(\lambda_j) - r_i(\lambda_{j-1}) \right]. \quad (4)$$

The distortion estimate quite successfully matches to the real distortion as shown in Fig. 5.

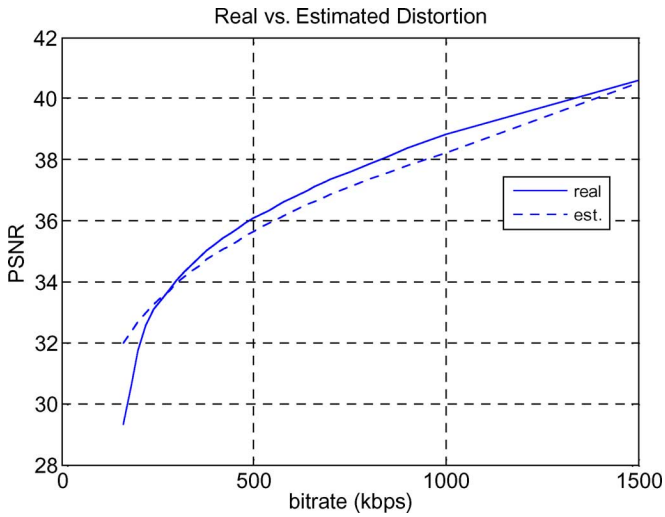


Fig. 5. Real versus estimated distortion from rate-distortion slopes and code-block lengths (in R-D hint tracks) for Foreman-CIF, $L = 40$ layers.

2) Determination of Design Parameters for a Special Case—An Example: A closed-form solution to the problem of determining the optimal design parameters in terms of rate-distortion slopes can be achieved for only special cases. Here, we provide an example in order to demonstrate the determination of design parameters for a special case of two peers available with identical bandwidth (R) and packet loss rates (p). Since there are only two paths, we set $N_B = 2$ and $N_E = 0$, i.e., no enhancement description is generated. Identical bandwidth and packet loss rates necessitate the use of balanced descriptions, which can be achieved by setting assignment vectors as $C_1^{B\prime} = [0,1]$ and $C_2^{B\prime} = [1,0]$. λ_L^B and λ_H^B can be found by a Lagrangian relaxation using packet loss rates and distortion expression. Expected distortion for a code-block i can be written as

$$D^i = (1-p)^2 D_1^i + p(1-p) D_2^i + p(1-p) D_3^i + p^2 D_4^i \quad (5)$$

where p is the packet loss probability and D_1^i, D_2^i, D_3^i and D_4^i are distortions, respectively, when i) both versions (high rate-low rate) of the code-block i arrive, ii) only low rate version arrives, iii) only high rate version arrives, and iv) none

of the code-blocks arrive. Here, we assume packet loss probability is equal to code-block loss probability, i.e., we assume code-block losses are also independent. The decoder uses the high rate coded code-block, if it exists; otherwise, it uses the low rate code-block. If none of the code-blocks are available, no concealment is performed. Hence

$$D_1^i = d_i\left(\lambda_H^B\right) \tag{6}$$
$$D_2^i = d_i\left(\lambda_H^B\right) \tag{7}$$
$$D_3^i = d_i\left(\lambda_L^B\right). \tag{8}$$

Bits spent on motion vectors, other overhead, and the distortion where no code-block is available can not be minimized; therefore, the relevant objective function to minimize becomes

$$D^i = (1-p)d_i\left(\lambda_H^B\right) + p(1-p)d_i\left(\lambda_L^B\right). \tag{9}$$

Since bit allocation for both high and low bitrate code-blocks are performed using EBCOT, we can safely assume that the total rate for one code-block is constant, i.e.,

$$r_i\left(\lambda_L^B\right) + r_i\left(\lambda_H^B\right) = r_i^T \tag{10}$$

where r_i^T is the rate of the code-block when all bits are spent on the high rate description.

From the minimization of the expression in (9), by setting the derivative to zero, $\partial D^i/\partial r_i(\lambda_L^B) = 0$, we get

$$\frac{\partial d_i\left(\lambda_H^B\right)}{\partial r_i\left(\lambda_L^B\right)} + p.\frac{\partial d_i\left(\lambda_L^B\right)}{\partial r_i\left(\lambda_L^B\right)} = 0. \tag{11}$$

Taking derivative of both sides of (10) with respect to $\partial r_i(\lambda_L^B)$, we reach $\partial r_i(\lambda_L^B)/\partial r_i(\lambda_H^B) = -1$. Replacing it in (11) and noting that $\partial d_i(\lambda_L^B)/\partial r_i(\lambda_L^B) = \lambda_L^B$ and $\partial d_i(\lambda_H^B)/\partial r_i(\lambda_H^B) = \lambda_H^B$, we get

$$\lambda_H^B = p.\lambda_L^B. \tag{12}$$

Hence, the optimum high and low slopes for code-blocks can be found by jointly iterating high/low rates ($r_i(\lambda_L^B)$, $r_i(\lambda_H^B)$) and slopes (λ_L^B, λ_H^B) to satisfy both (10) and (12). This procedure is analogous to bisection search methods used in finding the optimal rate-distortion slope that corresponds to a total rate in Lagrangian rate-distortion optimization [24].

NS-2 simulation results of the proposed derivation of high/low rates for this special case can be found in Section IV-D2. In the following section, we explain how to optimize the design parameters for adaptive P2P video streaming with arbitrary number of paths under different network conditions.

III. Adaptive P2P Video Streaming Using F-MDC

The P2P streaming session is initiated by the receiver with a *peer query process*. Peer query techniques, where the receiver finds all available peers that can serve the requested video (i.e., has the scalable coded video) are reported in [27]. The receiving peer then sends the total number of available sending peers, and which description (s) they should send to each available sending peer. Upon receipt of this information, all sending peers start sending their assigned descriptions, and the receiver starts the playback after the usual pre-roll delay. The receiver continuously monitors the quality of all paths from each sending peer through a *path measurement process*. The path measurement process, explained in detail in the next subsection, is used to estimate the TCP-friendly bandwidths of all paths from the sources to the receiver and packet loss correlation between these paths. The receiver performs rate-distortion optimization using the packet loss rates and the estimated TCP-friendly bandwidths obtained during the streaming of the previous GOP in order to determine the F-MDC design parameters explained in Section II-C for the current GOP. The initial parameter set at the beginning of the streaming session may depend only on the video content and resolution. Fig. 6 illustrates the overview of the proposed streaming system.

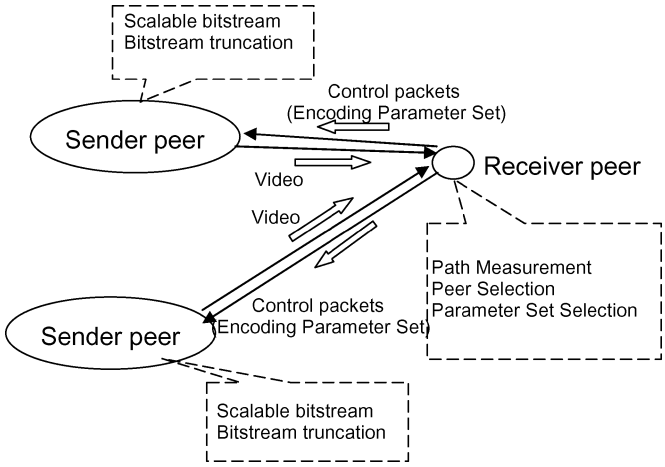


Fig. 6. Overview of the proposed P2P streaming system.

The proposed receiver-driven streaming algorithm is shown in Table II. One advantage of the proposed system is that it enables low pre-roll delay, since multiple descriptions generated with the initial F-MDC parameter set are received and displayed while the quality of each path is measured. The initial design parameter set can be determined according to expected network conditions and video content.

A. Path Measurements and Peer Selection

During a streaming session, the receiver measures the following parameters for each path m i) *average packet loss rate* (p_m), ii) *average receiving bandwidth* (R_m), and iii) *packet loss correlation between path* m *and path* k

The average packet loss rates p_m are computed using the techniques described in the Real-time Transport Protocol (RTP) [26]. Given the packet loss rate p_m, the average TFRC bandwidth R_m for m^{th} sending peer is estimated using the TFRC algorithm [29] as

$$R_m = \frac{S}{RTT_m\sqrt{\frac{2p_m}{3}} + t_{RTO_m}\left(3\sqrt{\frac{3p_m}{8}}\right)p_m\left(1+32p_m^2\right)} \tag{13}$$

where t_{RTO_m} denotes retransmit time-out value, RTT_m denotes the round trip time, and S denotes the packet size. It is assumed that $t_{RTO_m} = 4.RTT_m$.

TABLE II
P2P STREAMING ALGORITHM (AT THE RECEIVING PEER)

```
1.  Peer query, find all available (M) peers with encoded video
2.  For each peer m ; 1 ≤ m ≤ M ,
3.     Send initial, heuristically determined, design parameter set
4.           Measure path performance while receiving the video
5.              For each GOP
6.                Determine the design parameters through R-D optimization(see Section 3.3);
7.    For each peer m ; 1 ≤ m ≤ M ,
8.    Send the parameter set( λ_L^B*, λ_H^B*, λ_L^E*, λ_H^E*, λ_S*, N_B*, N_E*, C_1^B*, ..C_{N_B}^B*, C_1^E*, ..C_{N_E}^E* ) through control
    packets
```

We estimate packet loss correlations between multiple paths in order to determine if paths have shared links. We need to avoid receiving multiple base descriptions over paths with shared links, because there is no benefit to using MDC if a shared link is broken. To this effect, let $\varphi(m,k)$ denote the number of lost packets in paths m and k within the same time interval, and let $P(m,k)$ denote the total number of lost packets in paths m and k, and τ be a threshold value. If $\varphi(m,k)/P(m,k) \geq \tau$ then paths k and m are decided to share a common link; otherwise, paths k and m are assumed to be disjoint [27].

The receiver determines paths eligible to carry multiple base descriptions according to two criteria. Base layer MD paths should have available bandwidth R_m above some threshold (e.g., motion vector bitrate) and base layer MD paths should be disjoint. The receiver will not allocate two base descriptions on correlated paths.

B. Estimation of Total Distortion at the Receiver

Packet losses are typical in P2P networks due to unpredictable peer tune-outs and congestion caused by external traffic. In the Internet, packet losses due congestion are usually in bursts. Although the burst length affects observed distortion [28], in this work we assume that packet losses are independent for simplicity. Better models can also be used in our formulation [28]. The late packets are considered lost, and counted in the packet loss probability.

In order to estimate the distortion at the receiver, we restate the main decoding algorithm. In general, the decoder will receive more than one version of the code-block. In that case, the decoder will select the version at the highest rate among all versions of the code-block for reconstruction while discarding all other versions. In the worst case, when no version of the code-block is received, decoder will replace zeros in the place of the missing code-block. Note that, since code-blocks are in transform domain, the effect of a missing code-block does not show itself as a hole in the decoded video, but only shows as missing spatio-temporal frequency content. The expected distortion calculation and rate-distortion optimization process is GOP based. Both motion vectors and the lowest spatio-temporal frequency code-block should be received to decode the GOP. If motion vectors or lowest frequency code-block is not available, the decoder replaces the last frame of the previous GOP as error concealment mechanism. The decoding of enhancement descriptions depends on the availability of the code-block at the given starting layer (slope). If $\lambda_S = \lambda_L^B$, then the receiver is more likely to have the code-block at λ_S since any base description received is decodable at λ_L^B; hence, the enhancement descriptions are more likely to be decodable. However, if $\lambda_S = \lambda_H^B$, then the receiver is able to decode the code-block only when base code-block at λ_H^B is available, but higher quality is achieved if both enhancement and base version of the code-block are available. This tradeoff is captured in our expected distortion calculation and λ_S is chosen accordingly.

The expected distortion that we seek to minimize statistically measures the reconstructed quality at the receiver. If none of the base versions of the code-block is received, distortion incurred is d_i^0 for code-block i. Note that, if the lowest layer (slope) is chosen as minimum possible, this value of is d_i^0 available to the decoder as $d_i(\lambda_1)$, so we assume it is known to the receiver as $d_i^0 = d_i(\lambda_1)$.

Let $d_i(\lambda_j)$ be distortion reduction when i^{th} code-block truncated at layer j is received. Let us denote the probability of receiving i^{th} code-block at layer j is $p_i(\lambda_j)$. Then, the expected distortion for i^{th} code-block is

$$E\{d_i\} = \sum_{j=1}^{L} p_i(\lambda_j).d_i(\lambda_j). \tag{14}$$

Since all code-blocks are encoded independently, at the receiver side they can be decoded independently of other code-blocks. Let us compute the probabilities. First, we note that there can be five different values of nonzero probabilities of $\lambda_j : \{\lambda_1, \lambda_L^B, \lambda_H^B, \lambda_S + \lambda_L^E, \lambda_S + \lambda_H^E\}$. We will consider the two cases for λ_S (i.e., $\lambda_S = \lambda_L^B$ or $\lambda_S = \lambda_H^B$) separately in the optimization.

Let N_B the number of base descriptions, N_E be the number of enhancement descriptions where $N_E + N_B$ is constant and equal to available number of sending peers. Also, let $p_B^1, \ldots, p_B^{N_B}$ be the packet loss probabilities over the channels with rates $R_B^1, \ldots, R_B^{N_B}$ for base descriptions and $p_E^1, \ldots, p_E^{N_E}$ and $R_E^1, \ldots, R_E^{N_E}$ be packet loss probabilities and rates for enhancement descriptions. Given C_m^B and C_m^E the high and low rate coded code-blocks are known. Hence, let $p_{iBL}^1, \ldots, p_{iBL}^{N_{BL}}$, and $p_{iBH}^1, \ldots, p_{iBH}^{N_{BH}}$ $(N_{BL} + N_{BH} = N_B)$ be the packet loss probability of the channels over which high and low rate versions of the code-block i is sent. Let us first compute

$p_i(\lambda_1)$, which is the probability of receiving none of the base descriptions

$$p_i(\lambda_1) = \left(\prod_{k=1}^{N_{BL}} p_{iBL}^k\right) \cdot \left(\prod_{k=1}^{N_{BH}} p_{iBH}^k\right). \quad (15)$$

Now, let us compute $p_i(\lambda_L^B)$, which is receiving the base description truncated as slope λ_L^B. For this case, at least one of versions at λ_L^B should be received, and none of the versions at λ_H^B be received. Also, if $\lambda_S = \lambda_L^B$, none enhancement stream should be received to have the code-block at λ_L^B. Hence, the probability of reconstructing the code-block i at λ_L^B given $\lambda_S = \lambda_L^B$ is

$$p_i\left(\lambda_L^B | \left(\lambda_S = \lambda_L^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BL}} p_{iBL}^k\right) \cdot \left(\prod_{k=1}^{N_{BH}} p_{iBH}^k\right) \cdot \left(\prod_{k=1}^{N_{EL}} p_{iEL}^k\right) \cdot \left(\prod_{k=1}^{N_{EH}} p_{iEH}^k\right). \quad (16)$$

The probability of reconstructing the code-block i at λ_L^B given $\lambda_S = \lambda_H^B$ is

$$p_i\left(\lambda_L^B | \left(\lambda_S = \lambda_H^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BL}} p_{iBL}^k\right) \cdot \left(\prod_{k=1}^{N_{BH}} p_{iBH}^k\right) \quad (17)$$

since when $\lambda_S = \lambda_H^B$, it is sufficient to reconstruct code-block i at λ_L^B when at least one version at λ_L^B is received and none of the base descriptions at λ_H^B is received.

Similarly, probability of reconstructing the code-block i at λ_H^B given $\lambda_S = \lambda_L^B$ is

$$p_i\left(\lambda_H^B | \left(\lambda_S = \lambda_L^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BL}} p_{iBH}^k\right) \cdot \left(\prod_{k=1}^{N_{EL}} p_{iEL}^k\right) \cdot \left(\prod_{k=1}^{N_{EH}} p_{iEH}^k\right) \quad (18)$$

and reconstructing the code-block i at λ_H^B given $\lambda_S = \lambda_L^H$ is

$$p_i\left(\lambda_H^B | \left(\lambda_S = \lambda_H^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BL}} p_{iBH}^k\right) \cdot \left(\prod_{k=1}^{N_{EL}} p_{iEL}^k\right) \cdot \left(\prod_{k=1}^{N_{EH}} p_{iEH}^k\right). \quad (19)$$

To reconstruct the code-block at $\lambda_S + \lambda_L^E$ or $\lambda_S + \lambda_H^E$, first base version (at λ_S) should be available

$$p_i\left(\lambda_L^E + \lambda_S | \left(\lambda_S = \lambda_L^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BL}} p_{iBL}^k\right) \cdot \left(\prod_{k=1}^{N_{BH}} p_{iBH}^k\right) \cdot \left(1 - \prod_{k=1}^{N_{EL}} p_{iEL}^k\right) \cdot \left(\prod_{k=1}^{N_{EH}} p_{iEH}^k\right) \quad (20)$$

$$p_i\left(\lambda_L^E + \lambda_S | \left(\lambda_S = \lambda_H^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BH}} p_{iBH}^k\right) \cdot \left(1 - \prod_{k=1}^{N_{EL}} p_{iEL}^k\right) \cdot \left(\prod_{k=1}^{N_{EH}} p_{iEH}^k\right). \quad (21)$$

Similar to (20) and (21)

$$p_i\left(\lambda_H^E + \lambda_S | \left(\lambda_S = \lambda_L^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BL}} p_{iBL}^k\right) \cdot \left(1 - \prod_{k=1}^{N_{EH}} p_{iEH}^k\right) \quad (22)$$

$$p_i\left(\lambda_H^E + \lambda_S | \left(\lambda_S = \lambda_H^B\right)\right) = \left(1 - \prod_{k=1}^{N_{BH}} p_{iBH}^k\right) \cdot \left(1 - \prod_{k=1}^{N_{EH}} p_{iEH}^k\right). \quad (23)$$

However, when motion vectors and the lowest spatio-temporal frequency code-block are not received, video is not decodable for that GOP. Motion vectors are sent with every base description. The expected distortion of a GOP is

$$E\{D\} = \left(1 - \left(1 - \left(\prod_{k=1}^{N_{BL}} p_{1BL}^k \cdot \prod_{k=1}^{N_{BH}} p_{1BH}^k\right)\right)^{P_{MV}}\right) . D_0 + \left(1 - \left(\prod_{k=1}^{N_{BL}} p_{iBL}^k \cdot \prod_{k=1}^{N_{BH}} p_{iBH}^k\right)\right)^{P_{MV}} \cdot \sum_{i=1}^{C} \sum_{j=1}^{L} p_i(\lambda_j).d_i(\lambda_j) \quad (24)$$

where D_0 is the distortion estimate for that GOP when video is not decodable, P_{MV} denotes the number of packets that motion vectors and the lowest frequency subband code-block are included. All of those packets should be received to decode video. D_0 depends on the error concealment mechanism at the decoder. When motion vectors or the lowest spatio-temporal subband are not available, the bitstream is not decodable for that GOP. As error concealment, the last frame of the previous GOP is replicated for the whole GOP at the receiver. The encoder should repeat the concealment at the decoder to find D_0, which is simply the sum of mean square errors between every frame in the current GOP and the last frame of previous GOP. This information can be sent through control packets.

Note that, $\lambda = \{\lambda_1, \ldots, \lambda_L\}$ are already set before encoding and individual rates of code-blocks for each lambda $(r_i(\lambda_j))$ are found after encoding and attached to the packet headers in JPEG-2000. Also, packet loss rates for each GOP can be estimated for each channel and hence, from packet headers, distortion expression, (24), required in rate-distortion optimized rate allocation process, can be written for the whole GOP. For each GOP, a search over $\lambda = \{\lambda_1, \ldots, \lambda_L\}$, possible C_m^B, C_m^E and peers is performed to find the design parameters that minimize the expected distortion for the GOP.

C. Rate Distortion Optimization for Parameter Selection

This section discusses how F-MDC design parameters should be selected for network adaptation. Our goal is to find the best redundancy allocation given the R-D information for each code-block and channel conditions, subject to rate constraints on the paths. Formally, we write the optimization problem as follows

$$\begin{aligned}&\text{Find } \lambda_L^{B*},\lambda_H^{B*},\lambda_L^{E*},\lambda_H^{E*},\lambda_S^{*},N_B^{*},N_E^{*},C_1^{B*},..C_{N_B}^{B*},\\&\quad C_1^{E*},..C_{N_E}^{E*}=\arg\min E\{D\}\\&\text{subject to } \sum_{i=1}^{C} C_m^B(i).r_i\left(\lambda_H^B\right)+\left(1-C_m^B(i)\right).r_i\left(\lambda_L^B\right)\\&\quad + R_{MV}\le R_m^B,\quad 1\le m\le N_B \text{ and}\\&\sum_{i=1}^{C} C_m^E(i).\left(r_i\left(\lambda_H^E+\lambda_S\right)-r_i(\lambda_S)\right)+\left(1-C_m^E(i)\right)\\&\quad .\left(r_i\left(\lambda_L^E+\lambda_S\right)-r_i(\lambda_S)\right)\le R_m^E,\quad 1\le m\le N_E\end{aligned}\tag{25}$$

where R_{MV} denotes the motion vector and any other overhead bitrate. This optimization is performed for each GOP.

The receiver determines the design parameters for each peer based on the rate-distortion information for each code-block in a GOP. The optimization problem is NP-hard. Firstly, the optimal high/low slope (rate) code-block sets (C_m^B for $1 \le m \le N_B$, C_m^E for $1 \le m \le N_E$) should be determined along with the number of descriptions (N_B, N_E). This problem is referred to as set partitioning (or set covering) problem in integer programming and it is NP-hard in general [6], [31]. The full search over all possible sets can take significant complexity if truly optimal set separation is aimed; noting that there can be 2^C possible sets (high-low) for each GOP. However, we already constrain the possible sets into the ones that can be written in modulo form, sacrificing from the absolute optimality. To select the optimal combination of N_B, N_E, C_m^B, C_m^E, we perform full search over possible N_B, N_E, C_m^B, C_m^E values.

The optimization problem given in (25) is a constrained optimization problem with $N_B + N_E$ constraints when (C_m^B for $1 \le m \le N_B$, C_m^E for $1 \le m \le N_E$) are set. This problem is a linear combination of convex bit allocation problems; hence, it is a convex problem. This constrained problem can be converted to unconstrained optimization problem using a Lagrangian relaxation for each constraint, where γ_m^B is the Lagrange multiplier for rate constraint on m^{th} base channel, $(1 \le m \le N_B)$ and γ_m^E is Lagrange multiplier for rate constraint on m^{th} enhancement channel, $(1 \le m \le N_B)$

$$\begin{aligned}&\text{Find } \lambda_L^{B*},\lambda_H^{B*},\lambda_L^{E*},\lambda_H^{E*},\lambda_S^{*}=\arg\min J\\&J=E\{D\}\\&\quad+\sum_{m=1}^{N_B}\gamma_m^B\left[\sum_{i=1}^{C}\left(C_m^B(i).r_i\left(\lambda_H^B\right)+\left(1-C_m^B(i)\right).r_i\left(\lambda_L^B\right)\right)\right.\\&\qquad\left.-\left(R_m^B-R_{MV}\right)\right]\\&\quad+\sum_{m=1}^{N_E}\gamma_m^E\left[\sum_{i=1}^{C}C_m^E(i).\left(r_i\left(\lambda_H^E+\lambda_S\right)-r_i(\lambda_S)\right)+\left(1-C_m^E(i)\right)\right.\\&\qquad\left.\left(r_i\left(\lambda_L^E+\lambda_S\right)-r_i(\lambda_S)\right)-R_m^E\right].\end{aligned}\tag{26}$$

The problem is still too complex since the objective function J is not separable in terms of variables, e.g., $r_i(\lambda_H^E + \lambda_S)$ term has both λ_H^E and λ_S which can be λ_L^B or λ_H^B. To simplify the problem, we separate the problem into two steps, first, we find the optimal pair of λ_L^B, λ_H^B that minimizes the expected distortion related to the base descriptions. Then, for each case of λ_S ($\lambda_S = \lambda_L^B$ or $\lambda_S = \lambda_H^B$) optimal pair of λ_L^E, λ_H^E is found separately. That is

1)

$$\begin{aligned}\lambda_L^{B*},\lambda_H^{B*}=\arg\min&\sum_{i=1}^{C}\left(p_i\left(\lambda_L^B\right).d_i\left(\lambda_L^B\right)+p_i\left(\lambda_H^B\right).d_i\left(\lambda_H^B\right)\right)\\&+\sum_{m=1}^{N_B}\gamma_m^B\left[\sum_{i=1}^{C}\left(C_m^B(i).r_i\left(\lambda_H^B\right)+\left(1-C_m^B(i)\right)\right.\right.\\&\qquad\left.\left..r_i\left(\lambda_L^B\right)\right)\right].\end{aligned}\tag{27}$$

2)

i) Set $\lambda_S = \lambda_L^B$

$$\begin{aligned}\lambda_L^{E*},\lambda_H^{E*}=\arg\min&\sum_{i=1}^{C}\left(p_i\left(\lambda_L^E+\lambda_S\right).d_i\left(\lambda_L^E+\lambda_S\right)\right.\\&\qquad\left.+p_i\left(\lambda_H^E+\lambda_S\right).d_i\left(\lambda_L^E+\lambda_H\right)\right)\\&+\sum_{m=1}^{N_E}\gamma_m^E\left[\sum_{i=1}^{C}C_m^E(i).\left(r_i\left(\lambda_H^E+\lambda_S\right)\right)+\left(1-C_m^E(i)\right)\right.\\&\qquad\left.\left(r_i\left(\lambda_L^E+\lambda_S\right)\right)\right].\end{aligned}\tag{28}$$

ii) Set $\lambda_S = \lambda_H^B$ and repeat optimization in (28).

3) Choose λ_S, λ_L^{E*}, λ_H^{E*} that minimize $E\{D\}$

Now, let us focus on the first problem, given in (28). Karush–Kuhn–Tucker (KKT) optimality conditions [32] require that $\partial J/\partial r_i(\lambda_L^B) = 0$ and $\partial J/\partial r_i(\lambda_L^H) = 0$. The key observation in this problem is that only one of the constraints will be effectively active, i.e., some of the Lagrangian multipliers are zero and remaining ones are identical; hence, can be nested into an optimization problem with one Lagrangian multiplier γ_{act}. Now, let us find the relationship between the active Lagrange parameter γ_{act} and λ_L^{B*}, λ_H^{B*} using KKT conditions

$$\begin{aligned}\frac{\partial J}{\partial r_i\left(\lambda_L^B\right)}&=\sum_{i=1}^{C}\left(p_i\left(\lambda_L^B\right).\lambda_L^B+p_i\left(\lambda_H^B\right).\lambda_H^B.\frac{\partial r_i\left(\lambda_H^B\right)}{\partial r_i\left(\lambda_L^B\right)}\right)\\&\quad+\gamma_{act}^B\left(C_{act}^B(i).\frac{\partial r_i\left(\lambda_H^B\right)}{\partial r_i\left(\lambda_L^B\right)}+\left(1-C_{act}^B(i)\right)\right)\\&=0\end{aligned}\tag{29}$$

$$\sum_{i=1}^{C}\left(C_{act}^B(i).r_i\left(\lambda_H^B\right)+\left(1-C_{act}^B(i)\right).r_i\left(\lambda_L^B\right)\right)=R_{act}-R_{MV}.\tag{30}$$

Taking the derivative of both sides of (30) with respect to $r_i(\lambda_L^B)$, we obtain

$$\frac{\partial r_i\left(\lambda_H^B\right)}{\partial r_i\left(\lambda_L^B\right)}=\frac{\sum_{i=1}^{C}C_{act}^B(i)-C}{\sum_{i=1}^{C}C_{act}^B(i)}.\tag{31}$$

TABLE III
STREAMING SYSTEM ALGORITHM AT THE RECEIVER SIDE

```
1.  Find M paths eligible to carry base descriptions
2.      For each GOP
3.          For each possible C_m^B, C_m^E
4.              Find λ_L^B*, λ_H^B*, λ_L^E*, λ_H^E*, λ_S * with bisection search
5.                  Evaluate the total expected distortion via Eq.-24
6.      Find the optimal parameter set that result in minimum E{D}
```

TABLE IV
COMPARATIVE RESULTS PROPOSED/MD-MCTF (A) Foreman.qcif (ON THE LEFT) AND (B)Akiyo.qcif) ON THE RIGHT)

Redundancy	100kbps	200kbps	300kbps	Redundancy	50kbps	100kbps
20%	NA/27.4	31.64/30.72	34.75/33.94	20%	29.21/31.01	36.36/36.0
30%	29.44/28.72	34.03/33.42	37.02/36.43	30%	31.03/31.53	37.76/37.22
40%	31.09/30.53	34.99/33.67	38.01/37.54	40%	32.15/32.0	38.67/38.56
50%	32.35/32.12	35.76./34.2	38.59/38.11	50%	33.11/32.91	39.33/39.30

After substituting (31) into (28) and (29), and carrying out the algebra, we get

$$\frac{\lambda_H^B}{\lambda_L^B} = \frac{\sum_{i=1}^{C} p_i(\lambda_L)}{\sum_{i=1}^{C} p_i(\lambda_H)}. \tag{32}$$

λ_L^{B*} and λ_H^{B*} can be found by iterating (32) and rate constraint (30) with search methods such as bisection search [24]. Similar procedure can be used to find λ_L^{E*}, λ_H^{E*}. Experimental results confirm that this procedure results in the globally optimal point found by exhaustive search.

The pseudo-code of the full search optimized high/low layer selection and description/layer assignment method is shown in Table III.

D. On the Complexity of the Optimization

The complexity of the optimization increases linearly with the number of assignment vector choices and the number of description combinations. To decrease the complexity, we can develop suboptimal search algorithms such as searching around the optimal policy from the previous GOP. Also, iterative search algorithms, such as iterative sensitivity adjustment (ISA) used in rate-distortion optimal packet scheduling [25], can be used to decrease the complexity. Alternatively, a model-based approach (assuming a probability distribution for descriptions) can be used to achieve a closed form solution sacrificing from optimality. Since, the optimization is performed at the decoding side; it does not need cooperation of the sending peers which may not be willing to allocate computational resources. Also, we note that distortion estimation is model-based; in the sense that there is no need for multiple encode-decode type of rate or distortion estimation such as that of nonscalable coders such as H.264/AVC.

IV. RESULTS

A. Comparative Results on Compression Performance

The proposed method is compared to other multiple description coders with redundancy-distortion curves for some fixed rates. We used a wavelet coder based on JPEG-2000 4 [17], [18] with 3-level spatial and 4-level temporal decompositions.

Comparative results of the proposed coder and MD-MCTF (applied to the same scalable coder used) are provided in Table IV for case $N_B = 2$ descriptions ($N_E = 0$) at three different rates and four redundancy levels when only one description is received. Our coder outperforms MD-MCTF significantly at medium motion sequence (Foreman), however, for sequences with low motion (Akiyo) MD-MCTF performs comparable to our coder because MD-MCTF can properly estimate missing frames at sequences with low motion. Since MD-MCTF is reported to outperform MC-MDC [3] and we observed that our coder performs better than MD-MCTF in nearly all rate-redundancy levels, we did not compare our coder to MC-MDC coder.

B. P2P Streaming System Performance With Comparisons

For the streaming application, we simulate packet losses with an NS-2 simulator [30]. In Part-1, we simulate the proposed system for the example setting described in Section II-D to demonstrate that changing the redundancy of each description on the fly improves the streaming performance. Part-2 includes a comprehensive simulation of a P2P streaming system with TFRC rate allocation.

Part-1: In this part, we show the use of changing redundancy on the fly according to the derivation of the optimal high and low rates in Section II-D for the example basic scenario. This comparison shows the importance of the selection of high and low rates. The luminescence component of the Foreman sequence in QCIF format is coded with a wavelet coder with 3 spatial and 3 temporal decomposition levels for 296 frames at 30 fps. Other than the lowest frequency frame in the temporal decomposition, every frame is put into one packet with maximum packet

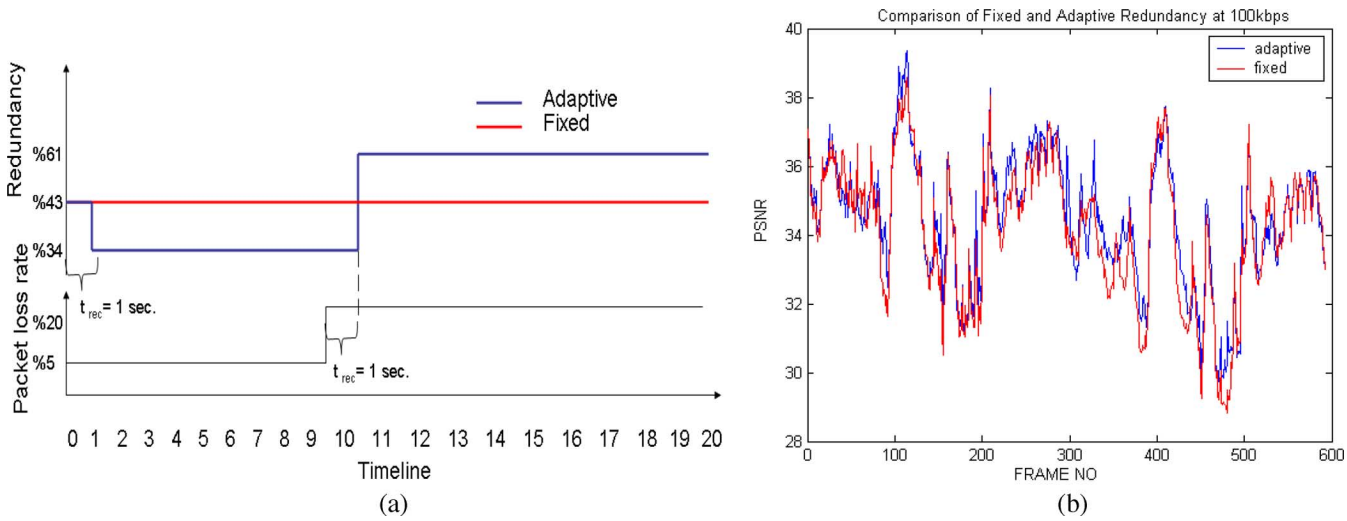


Fig. 7. (a) Loss rate and redundancy adaptation in time and (b) comparative result.

TABLE V
HIGH AND LOW RATE VALUES, CORRESPONDING TO HIGH/LOW SLOPES

	Packet Loss Rate=5%		Packet Loss Rate=20%	
	Low BR	High BR	Low BR	High BR
Adaptive	50 kbps	150 kbps	75 kbps	125 kbps
Fixed	60kbps	140 kbps	60kbps	140kbps

size of 1000 bytes. Every spatial resolution in the lowest frequency frame is put into one packet. All motion vectors for a GOP length 8 are put into a total of 2 packets. Traffic trace files generated in the coder are used in the ns-2 simulation to specify the timing and the size of each packet.

There are two senders who have the encoded video and description generator to generate description with any redundancy level. The last hop link is the bottleneck link with 100 kbps bandwidth and highest packet loss rate. Both senders send multiple descriptions over links with disjoint paths. Every path from each sender to receiver shares one link with 200 kbps bandwidth with external traffic. External cross traffic is randomly specified as 50% of the link capacity with exponentially distributed packet sizes and sending intervals. The simulation time corresponds to two full play time of the video segment, $\mathrm{T} = 20$ s. Sender's delay to detect a packet loss rate change, called the *loss detection delay*, is assumed to be $\mathrm{t_{rec}} = 1$ s. The rate of each description is set to the bottleneck bandwidth of $R = 100$ kbps.

The proposed system starts with medium redundancy (15%). After one loss detection delay period, $\mathrm{t_{rec}} = 1$ s, the sender adapts the redundancy level based on the packet loss rate 5%. At time $\mathrm{t} = 10.$ s, after a $\mathrm{t_{rec}} = 1$ s time from packet loss change, it changes the redundancy level according to the loss rate 20%. For comparison purposes, the performance of a test system with fixed level of redundancy is also simulated. Since the packet loss rate alternates between 5% and 20% during the simulation, the level of redundancy is fixed according to 15% loss rate. Fig. 7(a) depicts the change of the packet loss rate and the redundancy levels with respect to time for both cases. Packet loss rates are found by analyzing the ns-2 output trace files of the paths. The results are found by averaging 15 realizations of the simulation. Fig. 7(b) shows the PSNR values for every frame for both systems. It can be seen that the proposed system with adaptive redundancy level outperforms the fixed redundancy by 0.29 dB PSNR in the first half and 0.31 dB in the second half of the simulation. Table V shows the high and the low rates determined by our algorithm for the cases of 5% and 20% packet loss rates.

Part-2: In this part we simulate the proposed adaptive streaming system for a general network setting depicted in Fig. 8. The luminance component of the Foreman sequence in CIF format is coded with an embedded wavelet coder with 4 spatial and 4 temporal decomposition levels for 256 frames at 30 fps. We use fixed packet sizes as 500 bytes. For rate-distortion analysis we formed 20 layers between 150 kbps and 1 Gbps. We use TFRC rate control running at the receiver side. For the specific delay parameters used in the simulation, TFRC is in slow start in the first 10 GOP times ($\sim$5 s). We use this first 10 GOP for the path identification process, i.e., to determine if any path shares a common link with another or not. Since the TFRC rates are not sufficient even for sending the motion vector data in the first 5 GOP, only probe packets containing no information are sent during this period. From 5 GOP to 10 GOP time, we set the description number the same for each sender, independent of the TFRC and packet loss rates.

Fig. 8 illustrates the simulation setup where all paths share one link with external TCP connections. Paths 1 to 4 have 6 TCP connections that continuously send data during the simulation. Path 5 has 13 such TCP connections. Paths 1 to 4 have additional 9 TCP connections which start at random times after the first 10 GOP time and stop at 30 GOP time, and then start and stop periodically for 20 GOP times. We note that the bottleneck for

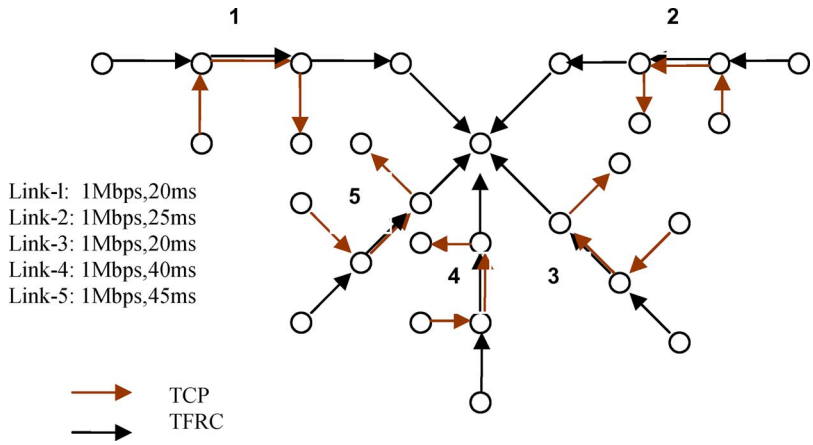


Fig. 8. Simulation setup.

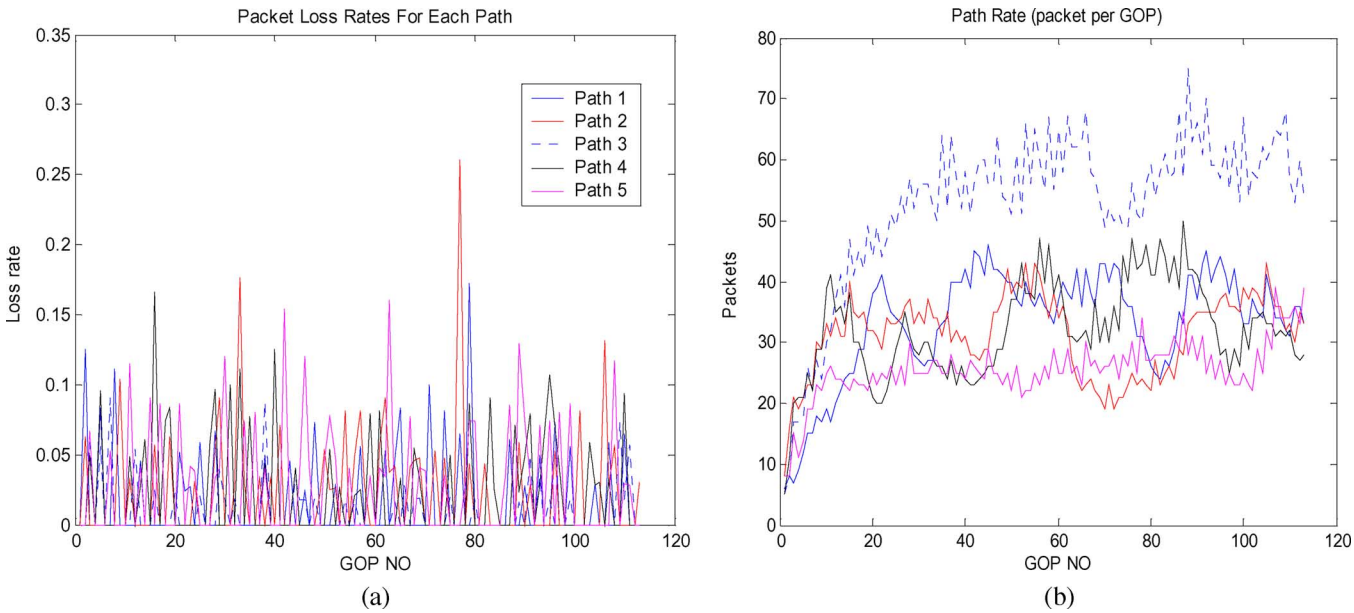


Fig. 9. (a) Loss rates as packets per GOP and (b) path rates as packets per GOP.

all paths is the shared link that carries both the TCP and the TFRC flows. All links have the same capacity as 1 Mbps and varying one way delays from 20 ms to 45 ms.

To compare, we simulated the performance of a fixed MDC streaming system where the number of descriptions is set to that of available paths (i.e., with no enhancement descriptions), the high/low rates are set to achieve the minimum redundancy, and code-block assignment vectors are set to $C'_1 = [1,0,0,0,0], C'_2 = [0,1,0,0,0], \ldots, C'_5 = [0,0,0,0,1]$. We note that from 5 GOP to 10 GOP time, compared and proposed systems send identical packets, hence have the same performance. Both the proposed and compared systems use TFRC rate control and hence, have the same rate and packet loss patterns as shown in Fig. 9. The proposed system however sends additional packets for rate-distortion hint tracks along with the motion vectors. The paths that carry the base descriptions also carry the rate-distortion hint tracks.

Fig. 10 shows that the proposed adaptive MDC streaming method outperforms fixed MDC by 1.3 dB in the average PSNR even when no peer tune outs or no significant throughput change occurs in any of the paths. As expected, the proposed adaptive system outperforms fixed MDC streaming by 2.5 dB for the scenario where peer-5 tune outs at time 20 s and peer-4 tunes outs at time 40 s. The fixed MDC streaming system always generates 5 descriptions with minimum redundancy level, matching the rates of the individual descriptions to the TFRC rates.

V. Conclusions

This paper presents two major contributions: 1) An F-MDC framework is introduced based on a highly scalable wavelet video codec with high compression efficiency. 2) A receiver-driven unicast (many to one) P2P streaming system using the novel F-MDC method is proposed. Optimal, in the rate-distortion sense, adaptation of the redundancy and rate of each description as well as the number of base and enhancement descriptions according to time-varying network conditions using the proposed hybrid layered-MD coder in the P2P streaming application scenario has been presented. The superiority of the

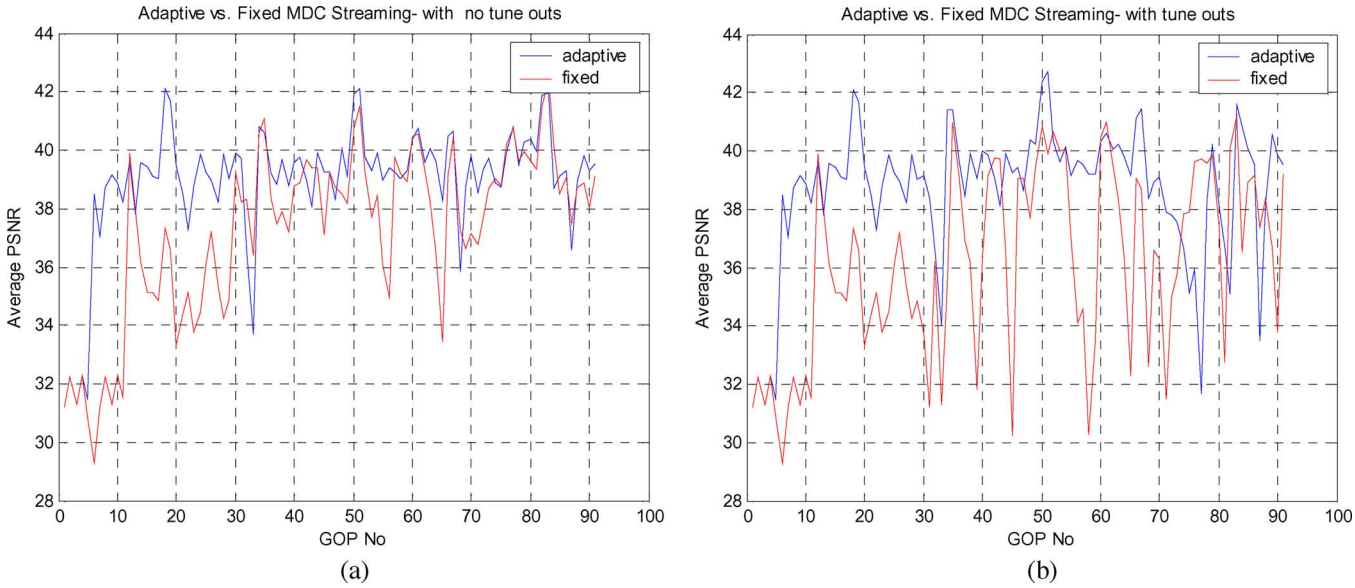


Fig. 10. Simulation results (a) with no tune outs and (b) with tune outs.

proposed adaptive system to fixed MDC is shown by means of NS-2 streaming simulations.

REFERENCES

[1] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, pp. 74–93, Sep. 2001.

[2] Y. Wang, A. Reibman, and S. Lin, "Multiple description coding for video communications," *Proc. IEEE*, vol. 93, pp. 57–70, Jan. 2005.

[3] Y. Wang and S. Lin, "Error resilient video coding using multiple description motion compensation," *IEEE Trans. Circuits, Syst., Video Technol.*, vol. 12, no. 6, pp. 438–452, Jun. 2002.

[4] D. Comás, R. Singh, A. Ortega, and F. Marqués, "Unbalanced multiple description video coding based on a rate-distortion optimization," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 1, pp. 81–90, Jan. 2003.

[5] M. Pereira, M. Antonini, and M. Barlaud, "Multiple description image and video coding for wireless channels," *Signal Process.: Image Commun., Special Issue on Recent Advances in Wireless Video*, vol. 18, no. 10, pp. 925–945, Nov. 2003.

[6] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1211–1225, 2003.

[7] M. van der Schaar and D. S. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003.

[8] R. Puri, K. Lee, K. Ramchandran, and V. Bharghavan, "Forward error correction (FEC) codes based multiple description coding for Internet video streaming and multicast," *Signal Process.: Image Commun.*, vol. 16, no. 8, pp. 745–762, May 2001.

[9] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Distributed video streaming using multiple description coding and unequal error protection," *IEEE Trans. Image Process.*, vol. 14, pp. 849–861, Jul. 2005.

[10] E. Akyol, M. Tekalp, and R. Civanlar, "Adaptive peer-to-peer video streaming with optimized flexible multiple description coding," in *Proc. IEEE Int. Conf. Image Processing*, 2006.

[11] E. Akyol, M. Tekalp, and R. Civanlar, "Optimal bit allocation in scalable multiple description video coding," in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*, 2005.

[12] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[13] J. Apostolopoulos, T. Wong, W. Tan, and S. Wee, "On multiple description streaming with content delivery networks," in *Proc. IEEE INFOCOM*, Jun. 2002.

[14] R. Rejaie and A. Ortega, "PALS: Peer-to-peer adaptive layered streaming," in *Proc. NOSSDAV'03*, Monterey, CA, Jun. 2003.

[15] X. Xu, Y. Wang, S. S. Panwar, and K. W. Ross, "A peer-to peer video-on-demand system using multiple description coding and server diversity," in *Proc. IEEE Int. Conf. Image Processing*, 2004.

[16] Y. Shen, Z. Liu, S. P. Panwar, K. W. Ross, and Y. Wang, "Streaming layered encoded video using peers," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, 2005.

[17] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform framework for highly scalable video compression," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

[18] D. Taubman, M. Reji, D. Maestroni, and S. Tubaro, SVC Core Experiment 1—Description of UNSW Contribution MPEG doc. m11441, Oct. 2004.

[19] S. Sen and J. Wang, "Analyzing peer-to-peer traffic across large networks," *IEEE/ACM Trans. Networking*, vol. 12, no. 2, Apr. 2004.

[20] T. Nguyen and A. Zakhor, "Multiple sender distributed video streaming," *IEEE Trans. Multimedia*, vol. 6, no. 2, Apr. 2004.

[21] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[22] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.

[23] D. S. Taubman and M. W. Marcellin, JPEG 2000: Image Compression Fundamentals, Standards, and Practice ser. Kluwer Engineering and Computer Science, 2002.

[24] A. Ortega and K. Ramchandran, "Rate-distortion techniques in image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[25] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 390–404, Apr. 2006.

[26] RFC 1889-RTP: A Transport Protocol for Real-Time Applications.

[27] O. Tickoo, S. Kalyanaraman, and J. Woods, "Efficient path aggregation and error control for video streaming," in *Proc. IEEE Int. Conf. Image Processing*, Singapore, 2004.

[28] Y. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Does burst length matter?," in *Proc. IEEE ICASSP*, Hong Kong, China, Apr. 2003.

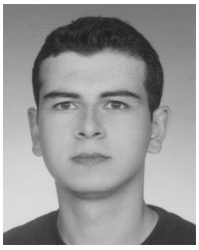
[29] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-based congestion control for unicast applications," in *Proc. ACM SIGCOMM*, Aug. 2000.

[30] The Network Simulator NS-2 [Online]. Available: http://www.isi.edu/nsnam/ns

[31] L. A. Wolsey and G. L. Nemhauser, *Integer and Combinatorial Optimization*. New York: Wiley, 1999.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[33] T. Tillo, M. Grangetto, and G. Olmo, "Multiple Descrition Image coding based on Langrangian rate allocation," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 673–683, Jan. 2007.



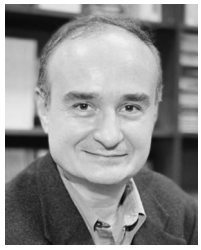
Emrah Akyol received the B.S. degree in electrical and electronics engineering in 2003 from Bilkent University, Ankara, Turkey, and the M.S. degree in electrical and computer engineering in 2005 Koç University, Istanbul, Turkey. He is currently pursuing the Ph.D. degree at the University of California, Los Angeles.

His research interests are in image-video compression and transmission. Between June and November 2006, he was a Research Intern at HP Labs, Palo Alto, CA, where he worked on complexity related problems of media compression. He has co-authored several publications and one pending patent application.

A. Murat Tekalp (F'05) received the Ph.D. degree in electrical, computer, and systems engineering from Rensselaer Polytechnic, Troy, NY, in 1984.

He was with Eastman Kodak Company, Rochester, NY, from 1984 to 1987, and with the University of Rochester from July 1987 to June 2005, where he was promoted to Distinguished University Professor. Since June 2001, he has been a Professor at Koç University, Istanbul, Turkey. His research interests includes video compression and streaming, motion-compensated video filtering for high-resolution, content-based video analysis and summarization, multicamera video processing, and protection of digital content. He authored the textbook *Digital Video Processing* (Englewood Cliffs, NJ: Prentice-Hall, 1995). He holds seven U.S. patents. His group contributed technology to the ISO/IEC MPEG-4 and MPEG-7 standards.

Dr. Tekalp was named as Distinguished Lecturer by IEEE Signal Processing Society in 1998. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992) and the IEEE TRANSACTIONS ON IMAGE PROCESSING (1994–1996). He has chaired the IEEE Signal Processing Society Technical Committee on Image and Multidimensional Signal Processing (Jan. 1996–Dec. 1997). He was appointed as the Technical Program Co-Chair for IEEE ICASSP 2000, and the General Chair of IEEE International Conference on Image Processing (ICIP) 2002. He is the Editor-in-Chief of the EURASIP journal *Signal Processing: Image Communication*.



M. Reha Civanlar (F'04) received the B.S. and M.S. degrees in electrical engineering from Middle East Technical University, Istanbul, Turkey, and the Ph.D. degree in electrical and computer engineering in 1984 from North Carolina State University, Raleigh.

He is Vice President and Media Lab Director for DoCoMo USA Labs, Palo Alto, CA. He was a Visiting Professor, Computer Engineering, Koç University, Istanbul, from 2002 to 2006. He also led a multinational research project on 3DTV transport. He is on the advisory boards of Argela Technologies and Layered Media, Inc. Before Koç, he headed the Visual Communications Research Department, AT&T Research. Prior to that, he was at Pixel Machines, Bell Laboratories. He has numerous publications, contributions to international standards, and over 40 patents. His research interests include networked video emphasizing the Internet and wireless networks and video coding.

Dr. Civanlar is recipient of a 1985 Senior Award of IEEE. He is a Fulbright Scholar and a Member of Sigma Xi. He served as an editor for IEEE TRANSACTIONS ON COMMUNICATIONS, TRANSACTIONS ON MULTIMEDIA, and the *Journal of Acoustics and Signal Processing*. He is currently an editor for *EURASIP Image Communications*. He served on MMSP and MDSP technical committees of the IEEE Signal Processing Society.